We are very grateful to the reviewers for their constructive and detailed feedback. We address specific reviewer comments and questions below; we will incorporate all the feedback into any final version of our paper.

Novelty of the paper (Reviewer 3): *What's the main novelty of this paper, in light of techniques in prior work?*

We agree that the core self-avoiding walk technique (as well as the color-coding technique) appeared in HS17 in the context of stochastic block models, and that nonbacktracking walks appeared in other prior work on stochastic block models. However, we feel that the main result of our paper – a polynomial-time algorithm for the spiked matrix model with heavy-tailed noise – is compelling on its own, because it provides a sharp algorithmic guarantee for a very simple and widely-studied problem. Our results also illustrate the versatility of the self-avoiding walk technique, in directions which were not obvious from prior literature, which focused on stochastic block models. For example,

- the technique can provide sharp guarantees for **general noise distributions**, not just **discrete distributions**,
- the technique does not require the assumption that the entries in **spiked vector** are sampled from some **i.i.d distribution with zero mean and unit variance**,
- the technique can extend beyond matrix settings, to handle **spiked tensor models**.

Furthermore, our work overcomes nontrivial technical difficulties in extending the self-avoiding walk method to these more general settings – for instance, allowing general spike vectors x (rather than $x \in \{\pm 1\}^n$ as in the block model setting) requires a substantially more challenging analysis of both the mean and the variance of the self-avoiding walk estimator.

Exposition: *Writing in sections 2 and 3 is uneven*

Several reviewers remarked that while the introduction to our paper is well written, the exposition in sections 2 and 3 is more uneven. We are grateful for this feedback. We have already worked to improve the exposition in these sections, and we will additionally incorporate the feedback from reviewers in the final version of our paper.

Technical clarifications and detailed responses

Dependence on δ (Reviewer 1): The dependence of δ in the main theorem is indeed $n^{\text{poly}(1/\delta)}$.

Clarifications in proof of Theorem 2.4 (Reviewer 3): We will include proof of the graph-theoretic relation $p \le r - s - k$ and the bound $n^{2(\ell-1)-r}\ell^{O(r-k)}$ in a revised full version. Following the line 227, the second expectation is uniformly taken over the labeling of $2(\ell-1)-r$ vertices in $\alpha \cup \beta$. Finally, '$\ell^{O(1)}$ diminish' means that for $r = k$, the bound is given by $(1+n^{-\Omega(1)})\lambda^{2\ell}n^{2(\ell-1)}n^{-k}\lambda^{-2k}$, without additional $\ell^{O(1)}$ factor. Again, we will clarify these points upon revision.

Clarification in the statement of lemma 2.5(Reviewer 3): Let $V \subseteq [n]$, $\|x\| = \sqrt{n}$ and $t_1, t_2 \in \mathbb{N}$. We define the quantity $S_{t_1,t_2,V} = \mathbb{E}_{(v_1,\ldots,v_{t_1+t_2})\subseteq[n]\setminus V}\left[\prod_{i=1}^{t_1} x_{v_i}^2 \prod_{i=t_1+1}^{t_1+t_2} x_{v_i}^4\right]$ where $(v_1, v_2, \ldots, v_{t_1+t_2})$ is uniformly sampled from all size-(t_1+t_2) ordered subsets of $[n] \setminus V$ (without repeating elements). Then assuming $|V|, t_1, t_2 = O(\log n)$ and $\|x\|_\infty^2 = n^{1-\Omega(1)}$, we have $S_{t_1,t_2,V} \le (1+n^{-\Omega(1)})\|x\|_\infty^{2t_2}$. Further if $t_2 = 0$, we have $S_{t_1,t_2,V} \ge 1 - n^{-\Omega(1)}$.

Notation in figures describing experiments (Reviewer 4): The label "naive" corresponds to the naive PCA algorithm(extracting the leading eigenvector). The label "worst" corresponds to the information-theoretically optimal recovery rate in case of Gaussian noise.

Why do some figures only have curves of truncation method and some figures only have curves of self-avoiding walk estimator? The scale 2000×2000 will be **computationally expensive** for **self-avoiding walk estimator**; at this scale we can still run the truncation-based algorithm. We thank reviewer 4 for the suggestion of putting all five methods in at least one figure (this can be done for matrices with smaller dimension) – we will add such a figure.

In the spiked matrix model, the authors only make assumption about the infinity norm of the spiked vector. However in the spiked tensor model, the authors assume that the entries are i.i.d sampled with zero mean and unit variance. Is this difference essential?(Reviewer 3, Reviewer 4): This difference is not essential. We can prove similar guarantee for spiked tensor model using nearly the same techniques, only assuming bound on the infinity norm of the spiked vector. However, the proof becomes lengthier. We will discuss this in a revised version of our paper.

Does the result in [PWBM18] require light tail? (Reviewer 4): [PWBM18] requires the 10-th moment of each entry in the noise matrix to remain constant as $n \to \infty$.

Typos: We are grateful to several reviewers for supplying a list of typos and small errors; we will address all of these in a revised manuscript.

Relevant literature: We are grateful to reviewer 4 for pointing out a relevant literature.